Knowledge Augmentation Services: Virtual Collections & Micro Articles

Kaylin Bugbee

2016 GHRC User Working Group Meeting
Sept 20-21, 2016

NASA

itsc
INFORMATION TECHNOLOGY AND SYSTEMS CENTER

THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Context

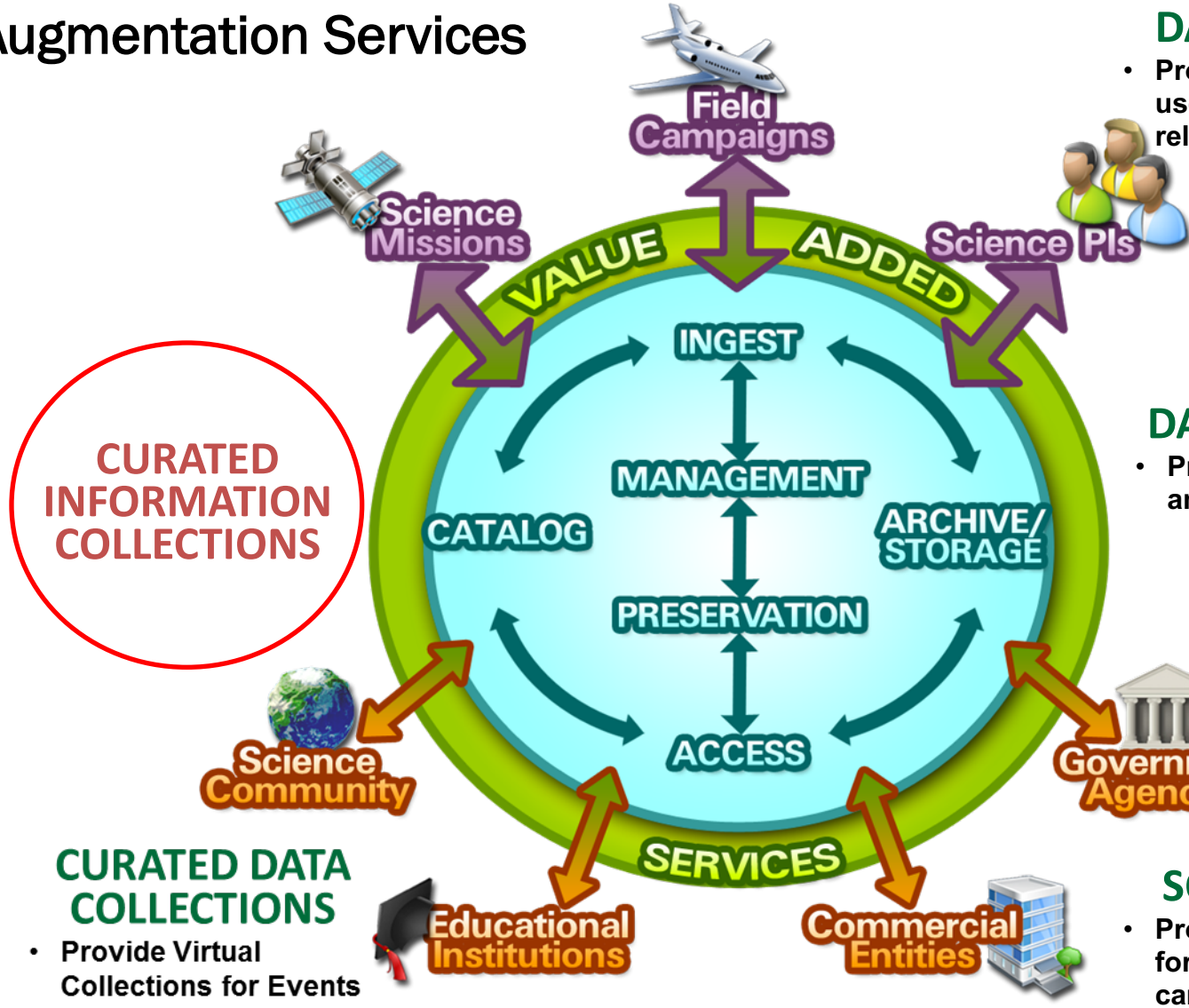
- Broadly speaking, there is a thirst for curated knowledge online
- Examples: Explainer YouTube video channels such as ASAP Science, Minutephysics, TedEd, etc…
- Users appreciate time and effort of putting concepts together by a trusted curator
- In the Earth sciences, “a much larger community of diverse users clamor to access, understand, and use climate data” (Overpeck et al., 2011)
- Initial user community
- Unanticipated users
- The curation and synthesis of Earth science data and related knowledge reduces effort and uncertainties for both the initial user community and unanticipated users and increases certainty of reward

Overpeck, J.T., Meehl, G.A., Bony, S., Easterling, D.R., Climate data Challenges in the 21st Century, *Science*, 33, DOI: 10.1126/science.1197869, 700 – 702, 2011.

Context

Knowledge Augmentation Services

DATA & SCIENCE EXPERTISE



CURATED INFORMATION COLLECTIONS

DATA DISCOVERY

- Provide tools that allow users to discover data and related information

DATA ACCESS

- Provide multiple methods to access the data

DATA EXPLORATION

- Provide tools to visualize and analyze the data

DATA USE

- Provide data recipes/code snippets to allow user to use the data

CURATED DATA COLLECTIONS

- Provide Virtual Collections for Events of Interest

SCIENCE PORTALS

- Provide customized portals for managing field campaigns and collecting data

Virtual Collections

Virtual Collections

What are virtual collections?

- Curation and special collection concept at libraries or museums
- In the digital environment, special collections have expanded and become more flexible
- Known as a virtual collection
- Curated around topic, organization or purpose
- Typically openly available
- Objects can be grouped in various way simultaneously
- The virtual collection is an Earth science data curation service that is growing in use
- Recommendation 17 from UWG requests “Create data bundles for scientists who want to study processes.”
- We define a virtual collection as an end product of a curation activity that searches, selects, and synthesizes diffuse data and information resources around a specific theme/topic or event

Virtual Collections

Use Case:

Created in collaboration with Dr. Patrick Gatlin (NASA MSFC)

Global Precipitation Measurement (GPM) mission Cold-season Precipitation Experiment (GCPEx) :

Synoptic scale snow event that occurred on February 24, 2012 across southern Ontario

Resource used to identify event:

Gail Skofronick-Jackson, David Hudak, Walter Petersen, Stephen W. Nesbitt, V. Chandrasekar, Stephen Durden, Kirstin J. Gleicher, Gwo-Jong Huang, Paul Joe, Pavlos Kollias, Kimberly A. Reed, Mathew R. Schwaller, Ronald Stewart, Simone Tanelli, Ali Tokay, James R. Wang, and Mengistu Wolde, 2015: Global Precipitation Measurement Cold Season Precipitation Experiment (GCPEX): For Measurement's Sake, Let It Snow. Bull. Amer. Meteor. Soc., 96, 1719–1741, doi: 10.1175/BAMS-D-13-00262.1.

Virtual Collections

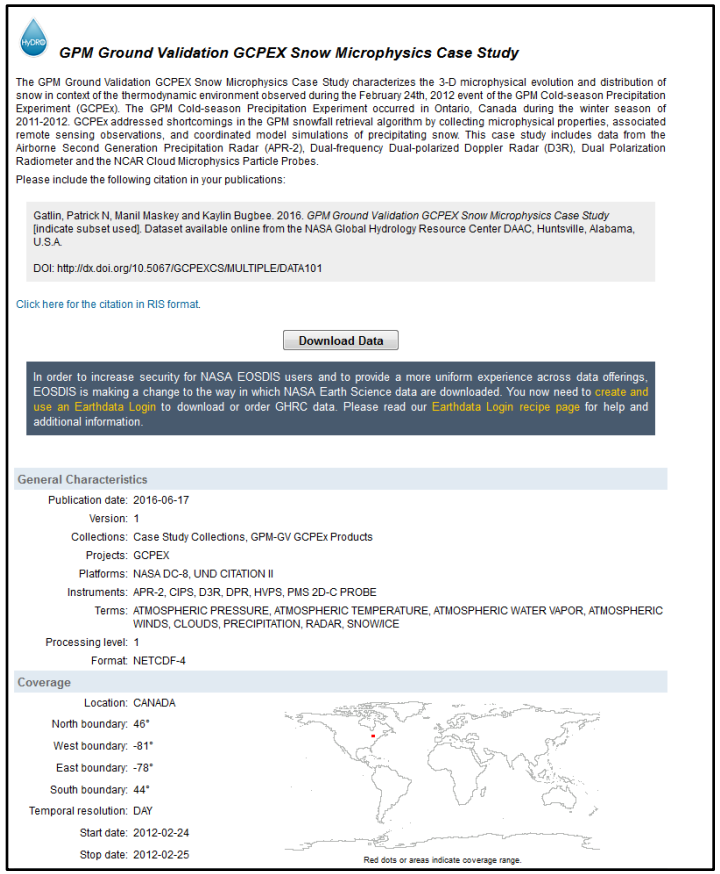
Use Case Data Search and Select Process -

1. *Formulate a science question*
   a. What is the 3-D structure of falling snow and how does its variability affect remotely sensed retrievals?
2. *What are the general keywords related to this science question that can be used to find the dataset required to help answer the question?*
   a. Remote sensing, snow
3. *What measurements are related to these keywords and what instruments take these measurements?*
   a. Microwave remote sensing
      a. Radar: D3R (ground), APR-2 (airborne)
      b. Radiometer: Dual polarization radiometer (ground)
   b. Snow particle size
      a. Cloud microphysics probes (airborne)
   c. Snow water equivalent
      a. Nevzorov probe (airborne)
4. *Which Intensive Observing Period (IOP) of the campaign was most of these measurements collected?*
   a. IOP must consist of a widespread and long duration snowfall event
   b. IOP must consist of measurements from at least ground and airborne platforms to address spatial variability
5. Data was selected from above instruments

Virtual Collections

Synthesis (Implementation):

1. iPython notebook
   a. Used an iPython notebook to create a virtual collection
   b. Application environment which allows interactive mode using a web browser
   c. Provides a distribution URL
   d. OPeNDAP based aggregation
2. Metadata
   a. Created test metadata record for this virtual collection in NASA's Common Metadata Repository (CMR)
3. Micro Article
   a. Short, interesting document that brings together data and key science concepts
   b. Describes golden case event and relevant data
4. GHRC Publication Workflow
   a. DOI (http://dx.doi.org/10.5067/GCPEXCS/MULTIPLE/DATA101)
   b. Called 'Case Study Collection' in Hydro



GPM Ground Validation GCPEX Snow Microphysics Case Study

The GPM Ground Validation GCPEX Snow Microphysics Case Study characterizes the 3-D microphysical evolution and distribution of snow in context of the thermodynamic environment observed during the February 24th, 2012 event of the GPM Cold-season Precipitation Experiment (GCPEx). The GPM Cold-season Precipitation Experiment occurred in Ontario, Canada during the winter season of 2011-2012. GCPEx addressed shortcomings in the GPM snowfall retrieval algorithm by collecting microphysical properties, associated remote sensing observations, and coordinated model simulations of precipitating snow. This case study includes data from the Airborne Second Generation Precipitation Radar (APR-2), Dual-frequency Dual-polarized Doppler Radar (D3R), Dual Polarization Radiometer and the NCAR Cloud Microphysics Particle Probes.

Please include the following citation in your publications:

Gatlin, Patrick N, Manil Maskey and Kaylin Bugbee. 2016. *GPM Ground Validation GCPEX Snow Microphysics Case Study* [indicate subset used]. Dataset available online from the NASA Global Hydrology Resource Center DAAC, Huntsville, Alabama, U.S.A.

DOI: http://dx.doi.org/10.5067/GCPEXCS/MULTIPLE/DATA101

Click here for the citation in RIS format.

Download Data

In order to increase security for NASA EOSDIS users and to provide a more uniform experience across data offerings, EOSDIS is making a change to the way in which NASA Earth Science data are downloaded. You now need to create and use an Earthdata Login to download or order GHRC data. Please read our Earthdata Login recipe page for help and additional information.

General Characteristics

Publication date:	2016-06-17
Version:	1
Collections:	Case Study Collections, GPM-GV GCPEx Products
Projects:	GCPEX
Platforms:	NASA DC-8, UND CITATION II
Instruments:	APR-2, CIPS, D3R, DPR, HVPS, PMS 2D-C PROBE
Terms:	ATMOSPHERIC PRESSURE, ATMOSPHERIC TEMPERATURE, ATMOSPHERIC WATER VAPOR, ATMOSPHERIC WINDS, CLOUDS, PRECIPITATION, RADAR, SNOW/ICE
Processing level:	1
Format:	NETCDF-4

Coverage

Location:	CANADA
North boundary:	46°
West boundary:	-81°
East boundary:	-78°
South boundary:	44°
Temporal resolution:	DAY
Start date:	2012-02-24
Stop date:	2012-02-25

Red dots or areas indicate coverage range.

Virtual Collections

Lessons Learned:

1. Search
   a. To identify a relevant event, re-visited publications, campaign blog, case study documents and other information relevant to GCPEx
   b. Science expert (Patrick) kept documentation of ‘golden cases’ during OLYMPEX campaign as the events took place to streamline the ‘select’ process at a later date when the data is published and readily available to start creating virtual collections
2. Select
   a. Metadata Quality
      a. Temporal
         i. Temporal information provided at collection level only - missing from granule level
         ii. Required manual selection of granules and subsetting to identified temporal period using OPeNDAP
      b. Spatial
         i. Spatial information provided at collection level only - missing from granule level
         ii. Required OPeNDAP to subset

Virtual Collections

Lessons Learned:

3. Synthesize
   a. Data Formats
      i. To use OPeNDAP, granules need to be provided in standard formats such as HDF or netCDF
      ii. Format translation code was required to convert non-standard granules (ASCII, XLS) to netCDF format
      iii. Note: OPeNDAP-based subsets of the data is only possible if there is a grid type defined for the parameter that we need to subset
   b. Processing Levels
      i. Granules with different processing levels were served under same dataset
      ii. Parameter-based subset could not be generalized for all the granules for a dataset with same spatial and temporal constraints
   c. System
      i. IPython notebooks could not be directly hosted within the DAAC due to security concerns. Therefore, only a static (already executed) version of the notebook is being hosted
   d. Other
      i. Data bundling, Science keyword/vocabulary, Metadata creation in CMR

Discussion

1. Do you think virtual collections/case study collections will make finding and using data easier?
2. Would you be interested in collaborating with the GHRC on creating new virtual collections?

NASA
itsc INFORMATION TECHNOLOGY AND SYSTEMS CENTER
THE UNIVERSITY OF ALABAMA IN HUNTSVILLE

Micro Articles

Micro Articles

What are micro articles?

- Micro articles (Tero et al., 2008) were originally defined as academic texts which are longer than a normal abstract but may be one tenth the size of a normal journal article.
  - Focus on gaps in the scientific research process that are generally not documented, published or shared.
- We have defined a micro article as a short, interesting document that brings together data and key science concepts. These documents are curated by both Earth and data scientists.
  - Short, 1 page 'Cheat Sheets' focused on our data holdings
- Micro article types at GHRC include:
  - Physical phenomena related to the data
  - Instrument used to collect the data
  - Event or Case Study (Golden Case)
  - Summary of a cornerstone publication concerning the data
- https://ghrc.nsstc.nasa.gov/home/micro-articles

Tero, Heiskanen, Kokkonen Juhana, A Hintikka Kari, Kola Petri, Hintsa Timo, N Pirjo, and kki. 2008. "Tutkimusparvi: The Open Research Swarm in Finland." *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*. Tampere, Finland: ACM. doi:http://doi.acm.org/10.1145/1457199.1457233.

Micro Articles

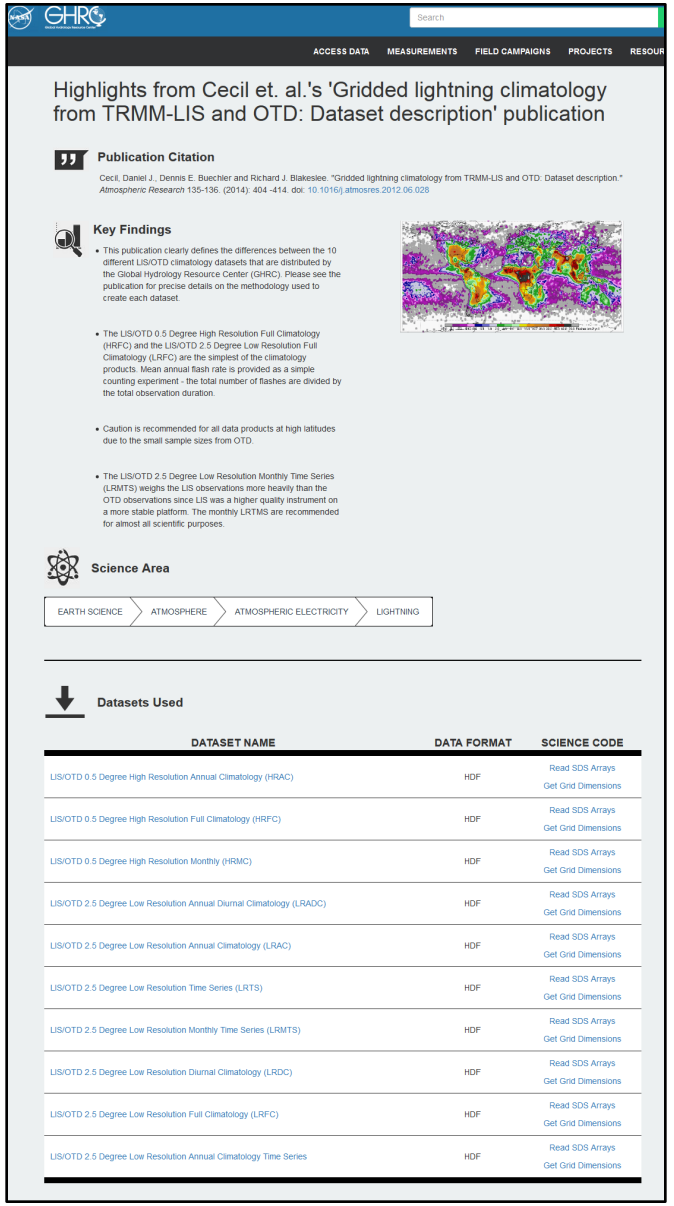
Micro Article – Event

- Key Information Included:
  - Event description
  - Description of science behind the event
  - Spatial and Temporal coverage
  - Access to virtual collection created for the event (DOI)
  - Data recipe
  - Information on datasets included in the virtual collection including:
  - Dataset access
  - Link to guide
  - Data format information
- URL: https://ghrc.nsstc.nasa.gov/home/micro-articles/snow-microphysics-event-during-gcpex-field-campaign#

Micro Articles

Micro Article – Publication

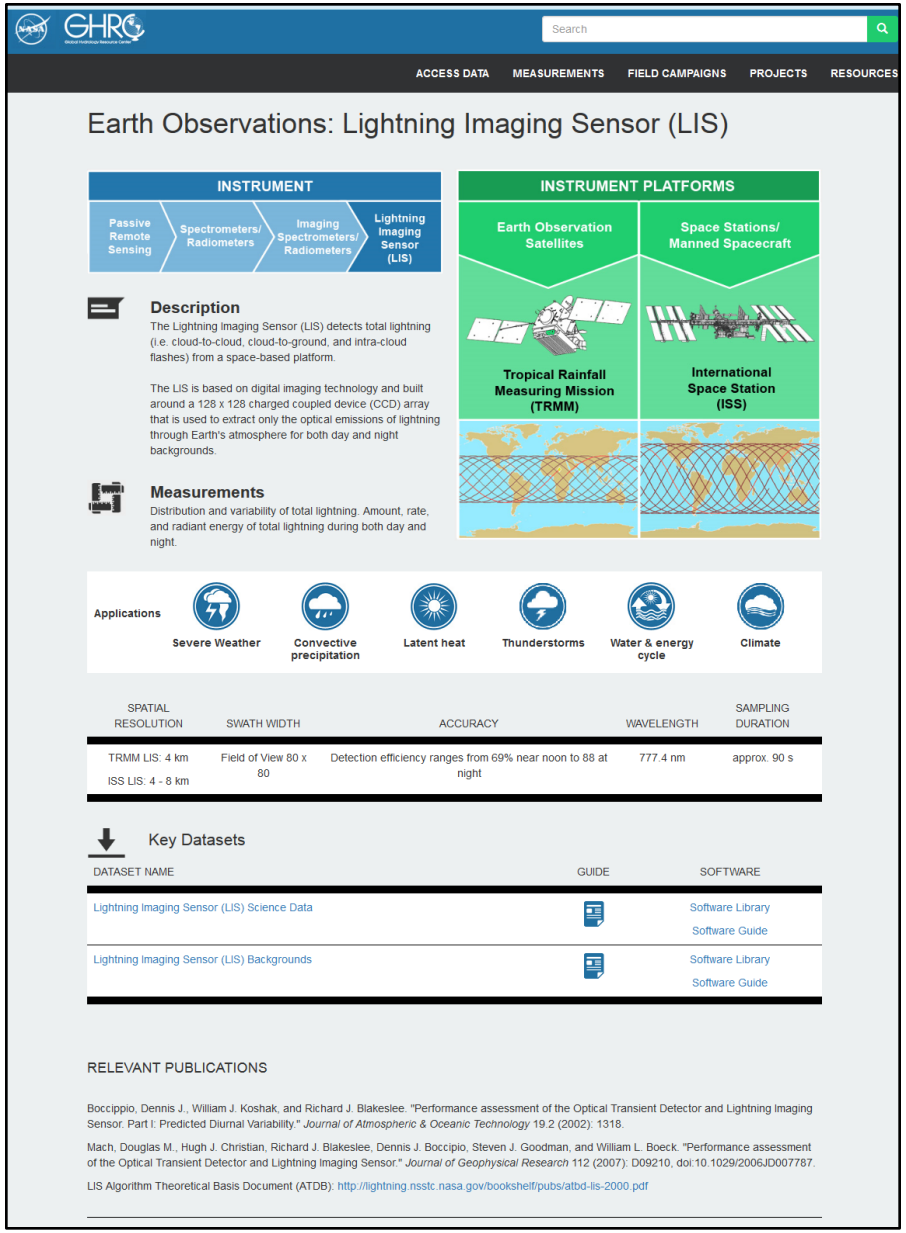
- Key Information Included:
  - Publication Citation and DOI
  - Key Findings from the Publication
  - Science Area (GCMD science keyword(s))
  - Datasets Used or Described in the Publication. Includes:
    - Link to dataset landing page
    - Data format
    - Data recipes (if applicable)
    - Science code
- URL: https://ghrc.nsstc.nasa.gov/home/micro-articles/highlights-cecil-et-als-gridded-lightning-climatology-trmm-lis-and-otd-dataset



Micro Articles

Micro Article – Instrument

- Key Information Included:
  - Instrument description
  - Platforms
  - Measurements
  - Applications
  - Spatial resolution
  - Swath width
  - Accuracy
  - Wavelength
  - Sampling duration
  - Key datasets
  - Relevant publications
- URL: https://ghrc.nsstc.nasa.gov/home/micro-articles/earth-observations-lightning-imaging-sensor



Micro Articles

Published

Weather Events

Snow Microphysics Event during GCPEx Field Campaign

Publication

Highlights from Cecil et. al.'s 'Gridded lightning climatology from TRMM-LIS and OTD: Dataset description' publication

Earth Observing Instruments

Lightning Imaging Sensor (LIS)

Upcoming

Phenomena (Type)

Lightning

Hurricanes

Publication

Highlights from Albrecht et al publication

Earth Observing Instruments

Disdrometers

Microwave radiometer

Discussion

1. Do you think micro articles are beneficial to the GHRC user community?
2. Are there other types of micro articles that you would like to see included?
3. Is there other key information that you would like to see included in a micro article?